



# Control of a PTZ camera in a hybrid vision system

François Rameau, Cédric Demonceaux, Désiré Sidibé, David Fofi

## ► To cite this version:

François Rameau, Cédric Demonceaux, Désiré Sidibé, David Fofi. Control of a PTZ camera in a hybrid vision system. International Conference on Computer Vision Theory and Applications, Jan 2014, Lisbon, Portugal. hal-00968425

**HAL Id: hal-00968425**

**<https://hal.science/hal-00968425>**

Submitted on 31 Mar 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Control of a PTZ camera in a hybrid vision system

François Rameau, Cédric Demonceaux, Désiré Sidibé and David Fofi

*Université de Bourgogne, Le2i UMR 6306 CNRS, 12 rue de la fonderie, 71200 Le Creusot, France.  
francois.rameau@u-bourgogne.fr*

**Keywords:** Fisheye camera, PTZ, target detection, hybrid vision system

**Abstract:** In this paper, we propose a new approach to steer a PTZ camera in the direction of a detected object visible from another fixed camera equipped with a fisheye lens. This heterogeneous association of two cameras having different characteristics is called a hybrid stereo-vision system. The presented method employs epipolar geometry in a smart way in order to reduce the range of search of the desired region of interest. Furthermore, we proposed a target recognition method designed to cope with the illumination problems, the distortion of the omnidirectional image and the inherent dissimilarity of resolution and color responses between both cameras. Experimental results with synthetic and real images show the robustness of the proposed method.

## 1 INTRODUCTION

Stereo-vision is one of the most explored topic in computer vision. The traditional approach is based on the use of two cameras of same nature mimicking the binocular human vision system (Marr and Poggio, 1977). Using two similar cameras drastically simplifies the steps of calibration and matching. However, in this paper we are dealing with a non standard stereo-vision rig, composed of a fisheye camera associated with a Pan-Tilt-Zoom (PTZ) camera. This kind of layout mixing different types of camera is called a hybrid vision system.

Omnidirectional cameras have the great advantage to provide a wide field of view (up to  $360^\circ$ ), however they often provide a limited and non-linear resolution. Furthermore, the use of omnidirectional sensors leads to a strong geometric distortion of the image, making most of the usual image processing methods inefficient.

On the other hand, a Pan-Tilt-Zoom camera is a zooming perspective camera which can be mechanically oriented in multiple directions. Despite a restricted field of view, the ability to steer the camera in the desired direction allows to cover a large region of the scene (up to  $360^\circ$  in panoramic and  $180^\circ$  in tilt direction, depending on the manufacturer). The zoom permits to obtain high resolution images of a specific region of interest (ROI). So, the versatility offered by those kind of camera is really appreciable for many applications especially in the field of video surveillance.

The couple composed of these two cameras combines the advantages given by both of them, that is to say a large field of view and an accurate vision of a particular ROI with an adjustable level of details using the zoom. Nevertheless, the control of the mechanical camera with information from the fisheye one is not straightforward. Moreover, the difference between the two cameras has to be taken into consideration. In fact, the model of projection as well as the color response and the resolution of the two sensors are greatly different.

Therefore, in this paper we propose a new approach able to cope with the previously mentioned problems and to find out the orientation of the PTZ camera for visualizing a defined ROI on the fisheye image.

The rest of the paper is organized as follows. First, we review previous methods using heterogeneous vision system. Section 2 is dedicated to an overview of the necessary background, containing a detailed description of the spherical model and of the epipolar geometry. The sections 3 and 4 respectively deal with the model of our system and its calibration. In Section 5, we describe our method of target detection through an hybrid stereo-vision system. Section 6 summarizes the results of our experiments. Section 7 concludes the paper.

### 1.1 Previous works

A hybrid vision system means that the two cameras of the rig have different characteristics. This combination allows to obtain additional information such

as an extension of the field of view, the extraction of 3D information or the study of a wider range of wavelengths (Cyganek and Gruszczynski, 2013). Recently, we noticed the emergence of composite stereo-vision sensors using the association of an omnidirectional and a perspective camera. This camera association is often considered as a bio-inspired method since the human retina can be divided into two parts - the foveal and the peripheral (Gould et al., 2007) - leading to a coarse vision in the peripheral region which is more sensitive to motion and a fine vision in the central part of the retina. In robotics, many articles have taken advantage of this specificity to facilitate the navigation using the wide angle camera, while the perspective camera allows to obtain accurate details of the environment. It is for instance the case in (Neves et al., 2008), where a soccer robot can navigate and detect the ball using an omnidirectional camera while a perspective camera is used for an accurate front view of the game. Similarly, in (Adorni et al., 2002) the authors proposed another approach for obstacle avoidance using merged information acquired from a PT and a catadioptric camera. Some innovative robotic applications using the previously described tandem of cameras do exist such as the robot photographer presented in (Bazin et al., 2011). Furthermore the use of this system is not only limited to ground robots but it is also applied to UAVs, for instance in (Eynard et al., 2012) the authors proposed to estimate the altitude and attitude of a UAV with a heterogeneous sensor.

However, the main application of hybrid vision remains the video surveillance because of the great versatility offered by those systems. (Ding et al., 2012) and (Puwein et al., 2012) are two representative examples of the possibilities offered by PTZ cameras network respectively for optimizing the surveyed area and for sport broadcasting. Many others papers are dealing with collaborative tracking and recognition using these types of cameras (Micheloni et al., 2010; Raj et al., 2013; Amine Iraqui et al., 2010). For video surveillance applications we often assume a static location of the rig, making possible to calibrate and use the hybrid stereo-vision system based on different a-priori. For instance in (Chen et al., 2008; Cui et al., 1998; Scotti et al., 2005), the planarity of the ground, the height of the devices, a size of a person or the alignment of the vertical axis of the cameras are supposed to be known. Another very usual approach is to create a look up table between the PTZ setpoints and the coordinates of the omnidirectional camera (Badri et al., 2007; Liao and Cho, 2008). This method of calibration assumes a fixed environment and required a cumbersome step of calibration. In this paper we pro-

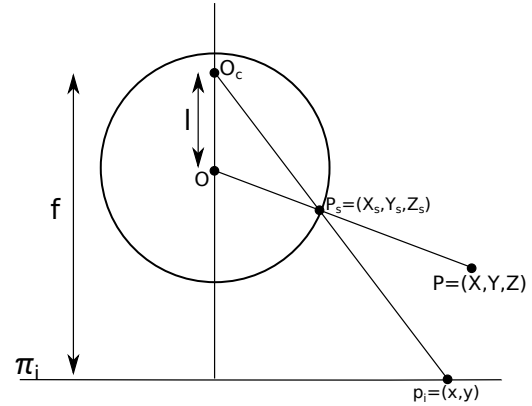


Figure 1: Unified spherical model

pose a more flexible method able to steer a rotating camera in the direction of a selected target from the wide angle image in an unknown environment.

## 2 Background

### 2.1 The unified spherical model

The cameras in our system have different projection models, it is possible to homogenize it using the unified spherical model defined by Barreto *et al.* (Barreto and Araujo, 2001) which remains valid for fisheye and perspective cameras (Ying and Hu, 2004). Theoretically, this model can only fit with SVP (Single View Point) cameras which is not the case of fisheye sensors. However it has been proved that this approximation still holds (Courbon et al., 2012).

Furthermore, it is also a suitable model for PT/PTZ cameras since the translation leads by the mechanical motions of the camera can be neglected (Rameau et al., 2012). Consequently the SVP assumption is also satisfied.

For any central camera, the image formation process can be described by a double projection on a Gaussian sphere (as shown in fig.1). First, a world point  $P$  is projected onto the sphere at the point  $P_s$ . Then, this first projection is followed by a second one on the image plane  $\pi_i$  inducing the pixel  $p_i$ . This projection starts from a point  $O_c$  located above the center of the sphere. The distance  $l$  between the point  $O_c$  and the center of the sphere  $O$  models the inherent radial distortion of the camera. This distance is null if we consider a perspective camera without distortion while  $l > 1$  for fisheye lens (Ying and Hu, 2004).

In this work we mainly use the inverse projection to back-project image plane's pixel on its equivalent unitary sphere. Basically this back-projection allows to

take the non-linear resolution as well as the camera's distortion into consideration.

A prior knowledge on the intrinsic parameters of the

camera  $K = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$  is necessary in order to

back-project a pixel  $p_i(x, y)$  onto a point lying on the sphere  $P(X_s, Y_s, Z_s)$ . Knowing those parameters, the projection can be expressed under the following form:

$$\begin{cases} Z_s = \frac{-2.l.\omega + \sqrt{(2.l.\omega)^2 - 4(\omega+1).(l^2.\omega-1)}}{2(\omega+1)} \\ X_s = x_t(Z_s + l) \\ Y_s = y_t(Z_s + l) \end{cases},$$

with  $\begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} \simeq K^{-1} p_i$  and  $\omega = x_t^2 + y_t^2$ .

## 2.2 Epipolar geometry

The epipolar geometry is the mathematical model relating two images of the same scene taken from different viewpoints. This well known geometry is based on the intersection of the image planes with the epipolar plane  $\pi_e$ . This plane is formed by the optical centers of the cameras and a 3D point  $X$  projected on both images in  $x$  and  $x'$ . The epipolar geometry can be mathematically formalized using the fundamental matrix  $F = K_2^{-T} T_{[x]} R K_1^{-1} = K_2^{-T} E K_1^{-1}$  (with  $K_1$  and  $K_2$  the intrinsic parameters of the cameras and  $E$  the essential matrix). Therefore the fundamental matrix links two corresponding points by the relation  $x'^T F x = 0$ .

The epipolar geometry remains valid in the context of an omnidirectional or a hybrid stereo-vision system (Fujiki et al., 2007). In fact, since we use the spherical representation of images the projective geometry is valid. In this configuration the centers of the spheres  $O_o$  and  $O_p$  are respectively the optical center of the omnidirectional camera and of the perspective camera. Thus, the baseline between cameras intersects each sphere in two positions, forming four epipoles  $e_1, e_2, e'_1$  and  $e'_2$  (see fig.2). Because the projective geometry is preserved the epipolar relation between points on the spheres can be expressed as follows:

$$P^T E P = 0, \quad (1)$$

where  $E = [t]_{\times} R$ , with  $t$  and  $R$  the translation and the rotation between the cameras. Hence, any selected point  $P$  on  $S_o$  defines a great circle  $C$  on  $S_p$ .

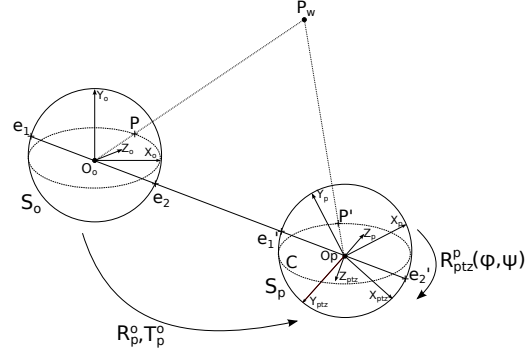


Figure 2: Model of our hybrid stereo-vision system

$S_o$	Spherical model of the fisheye camera
$S_p$	Spherical model of the PTZ camera
$O_o$	Center of $S_o$ , world coordinate system
$O_p$	Center of $S_p$ , $O_p = T_p^o$
$P$	Target point $\in S_o$
$P_w$	3D location of the target in the world
$E$	Essential matrix $E = T_{[x]} R$
$\pi_e$	Epipolar plane defined by $EP$
$C$	Epipolar great circle $\in S_p$
$P'$	Desired point $\in C$
$\Psi$	Angular setpoint of the camera in tilt
$\Phi$	Angular setpoint of the camera in pan
$K_p$	Intrinsic parameters of the PTZ camera
$K_o$	Intrinsic parameters of the fisheye camera
$T_p^o$	Translation between cameras
$R_p^o$	Rotation between the cameras for $\Psi = \Phi = 0$
$R_{ptz}^p(\Phi, \Psi)$	Rotation of the PTZ in its coordinate system
$(\vec{X}_o, \vec{Y}_o, \vec{Z}_o)$	Fisheye coordinate system
$(\vec{X}_{ptz}, \vec{Y}_{ptz}, \vec{Z}_{ptz})$	PTZ coordinate system
$(\vec{X}_p, \vec{Y}_p, \vec{Z}_p)$	Intermediate coordinate system for $R_{ptz}^p = I$

Table 1: Notations

## 3 Model of the system

Figure 2 summarizes all the possible geometric relationships between the 2 cameras and table 1 defines the notations used. For convenience, the fisheye coordinate system  $(\vec{X}_o, \vec{Y}_o, \vec{Z}_o)$  is taken as the reference of our model. Then, the position and orientation of the PTZ camera in its own coordinate system  $(\vec{X}_{ptz}, \vec{Y}_{ptz}, \vec{Z}_{ptz})$  is expressed with respect to the global coordinate system.  $(\vec{X}_{ptz}, \vec{Y}_{ptz}, \vec{Z}_{ptz})$  is the coordinate system of the PTZ camera for its calibrated position. Note that a translation  $T_{ptz}^p$  practically exist, it is the residual translation leads by the mechanisms used to rotate the camera.

Consequently, a point  $P_{ptz}$  in the PTZ coordinate system can be expressed in the global coordinate system as follow:

$$P_o = R_p^{oT} R_{ptz}^p(\Phi, \Psi)^T (P_{ptz} - T_{ptz}^p) - T_p^o, \quad (2)$$

where  $P_o$  is the point  $P_{ptz}$  in the omnidirectional coordinate system. In our case we consider the translation  $T_{ptz}^p$  as negligible as it is often the case (Agapito et al.,



2001), hence:

$$P_o = R_p^{oT} R_{ptz}^p(\phi, \psi)^T P_{ptz} - T_p^o. \quad (3)$$

## 4 Calibration of the hybrid stereo-vision system

### 4.1 Intrinsic calibration

As mentioned in the section 2, our cameras have to be calibrated individually in order to use the spherical model. For the PTZ camera (for a given level of zoom) we used the toolbox developed by Bouguet (Bouguet, 2008) in order to obtain the calibration matrix  $K_p$  while the fisheye camera has been calibrated using (Mei and Rives, 2007) giving  $K_o$  and  $l$ .

### 4.2 Extrinsic calibration

To determine the extrinsic parameters of the stereo rig at its initial position ( $R_p^o$  and  $T_p^o$  with  $R_{ptz}^p(0, 0) = I$ ), we propose to compute the homography  $H$  induced by a plane projected on our two spheres (Mei et al., 2008):

$$H \sim R_p^o - \frac{T_p^o n^T}{d}, \quad (4)$$

where  $\sim$  denotes the equality up to scale, with  $n^T/d$  the ratio between the normal of the plane and its distance to the optical center of the fisheye camera.

In order to compute this homography we use  $m$  points on a plane visible simultaneously by both cameras. Let  $P_o^i$  be the  $i^{th}$  point lying on the fisheye sphere  $S_o$  and  $P_p^i$  the  $i^{th}$  point on the PTZ's sphere  $S_p$ . The homography of the plane between the two views leads to the following relationship:

$$P_o^i \sim H P_p^i. \quad (5)$$

After computing  $H$ , we can extract  $R_p^o$  and  $T_p^o$  as an initialization for a non-linear refining method by solving the following optimisation problem:

$$\begin{aligned} \{R^*, T^*\} = \\ \underset{R, T}{\operatorname{argmin}} \sum_{i=1}^k [d^2(P_p^i, T_{p[\times]}^o R_p^o P_o^i) + d^2(P_p^{iT} T_{p[\times]}^o R_p^o P_o^i)], \\ \text{such that } R^T = R^{-1} \end{aligned}$$

with  $k$  the number of matched points,  $P_p^i, P_o^i$  the  $i^{th}$  matched points of the scene and  $d$  the geodesic distance between the point and its corresponding epipolar great circle.

## 5 Methodology

We want to find the angular setpoint of the PTZ camera ( $\psi, \phi$ ) in order to visualize a ROI selected on the fisheye camera. The proposed method can be decomposed into two main steps. The scanning of the epipolar great circle using the rotations of the PTZ camera followed by the detection of the region of interest.

### 5.1 Epipolar circle scanning

With a fixed sensor and without any a priori knowledge of the scene it is impossible to directly steer the PTZ camera in the desired direction. However, considering the centroid of the ROI ( $P_o^c$ ) on  $S_o$ , we can limit the search space to the set of points  $P_p^c \in S_p$  satisfying  $P_p^{cT} T_{p[\times]}^o R_p^o P_o^c = 0$ . This set of points draws a great circle  $C$  on  $S_p$ . We propose to scan this circle aligning the central axis of the PTZ camera on  $C$ .

Getting the setpoints of the camera to scan the circle is straightforward. Indeed, we obtain them by converting all points lying on  $C$  into spherical coordinates:

$$\begin{aligned} \forall P(X, Y, Z) \in S_p, \\ \begin{cases} \phi = \arccos(Z / \sqrt{X^2 + Y^2 + Z^2}) \\ \psi = \arctan(Y/X) \end{cases} \end{aligned}$$

Finally the rotations satisfying the following statement allow to center the PTZ camera on the circle:

$$N \cdot R_p^o R_{ptz}^p(\phi, \psi) [0 \ 0 \ 1]^T = 0,$$

with  $N = T_{p[\times]}^o R_p^o P_o^c$  the normal of the epipolar plane.

### 5.2 Detection of the ROI

This step aims to match a selected template on the fisheye image to its corresponding region through a series of perspective images. The template matching is one of the most challenging task in the context of a hybrid vision system because of the multiple problems already mentioned in the introduction of this article. Thus, in this section we propose an approach able to deal with the high dissimilarities between images. As we scan the epipolar circle with the PTZ camera, we get a set of images (an image for each position  $(\phi, \psi)$  of the camera). The detection task is thus to localize the target through the images acquired during the scanning step, and to find the corresponding setpoint to steer the camera to this specific ROI. First we select a patch on the fisheye image representing the ROI. Thereafter, we detect feature points on the patch as well as on the set of perspective images. Let  $p_o$ , and  $p_p$  be the Harris points detected on the fisheye and perspective images respectively. The

detected points are back-projected on the sphere, giving  $P_o$  and  $P_p$ . Most of these detected points can be rejected using the epipolar constraint  $|P_p^T E P_o| < \varepsilon$ , where  $\varepsilon$  is an arbitrary threshold. Then only the potential corresponding points are preserved. Note that this approach is purely geometric, which means that no photometric descriptors are used, making our matching robust to illumination changes, rotation and scale. Now, considering that the ROI is locally planar it is possible to compute the homography fitting the model and reject outliers simultaneously, using a RANSAC procedure (Fischler and Bolles, 1981). A homography is typically computed using 4 points, yet in the present case, since we know the rotation and translation between the cameras, we can accordingly reduce the number of points needed to compute the homography matrix  $H$  by pre-rotating the points  $P_p^i$  on the sphere  $S_p$ . Reducing the number of points needed to compute  $H$  allows to exponentially decrease the complexity of the method. Without loss of generality, the equation (4) can then be rewritten as follows:

$$H \sim I - \frac{T_p^o n^T}{d}. \quad (6)$$

Knowing  $T_p^o$ , the number of degrees of freedom of  $H$  is reduced to 3 which are the 3 entries of  $N_d = \frac{n^T}{d}$ .  $H$  is then expressed by:

$$H = \begin{bmatrix} 1 - \frac{n_x}{d} t_x & -\frac{n_y}{d} t_x & -\frac{n_z}{d} t_x \\ -\frac{n_x}{d} t_y & 1 - \frac{n_y}{d} t_y & -\frac{n_z}{d} t_y \\ -\frac{n_x}{d} t_z & -\frac{n_y}{d} t_z & 1 - \frac{n_z}{d} t_z \end{bmatrix}, \quad (7)$$

where  $T_p^o = [t_x \ t_y \ t_z]^T$  and  $N_d = [n_x \ n_y \ n_z]^T / d$ . To find out the 3 entries of  $N_d$  we solved  $P_p \times P_o H = 0$ . Every single point correspondences  $P_o(x, y, z)$  and  $P_p(x', y', z')$  gives 3 equations (of the form  $AN_d = b$ ) all linearly dependent to the following equation:

$$[t_y x z' - t_z x y' \quad t_y y z' - t_z y y' \quad t_y z z' - t_z y' z] N_d = y' z - y z', \quad (8)$$

so 3 points correspondences are enough to solve  $N_d$  using a singular value decomposition. The homography is finally computed using the equation (6). The output of the RANSAC process subsequently gives the inliers points and the optimal matrix  $H$  fitting our model.

To summarize the RANSAC algorithm applied here: 3 points are randomly selected among the points  $P_o$  and  $P_p$  to compute a homography matrix  $H$ , then the number of inliers ( $N_I$ ) is computed by,

$$k_i = \begin{cases} 1 & \text{if } \|P_o^i - H P_p^i\|^2 + \|P_p^i - H^{-1} P_o^i\|^2 < \tau \\ 0 & \text{else} \end{cases},$$

$$N_I = \sum_{i=1}^n k_i,$$



Figure 3: Hybrid stereo vision system used for the experiments

where  $\tau$  is an arbitrary threshold and  $n$  the total number of potential matched points. Afterwards, we reiterated the process until we get the homography fitting the largest number of inliers.

Finally, the resulting bounding box on the PTZ image can be easily drawn by transforming the coordinates of the corners of the selected ROI using  $H$ . The set-point to steer the camera in the desired direction is given by the spherical coordinates of the centroid of the patch  $P_p^c = H P_o^c$ .

## 6 Results

This section is dedicated to the assessment of our method, to do so we present here a quantitative analysis tested in a photo-realistic environment generated with a ray tracer software. A series of qualitative experiments in real conditions is also conducted to prove the robustness of our approach in such complex configurations.

In order to test our algorithm we have decided to use PovRay (Persistence of Vision Ray Tracer) for synthesizing fisheye and perspective views in a totally controlled environment. In fact, the internal parameters as well as the localization of the cameras are known and tunable. This assessment approach allows to work in a quasi-real scene where all the 3D information are known making the experiments as close as possible from the reality. In this series of tests, both camera have a  $640 \times 480$  pixels sensor.

We also conducted multiple experiments with real images where the results have been obtained using a fixed  $640 \times 480$ p's camera equipped with a fisheye lens to provide a field of view of  $180^\circ$ . The PTZ camera used is an AXIS 2130R which has the ability to rotate in pan and tilt directions respectively up to  $338^\circ$  and  $90^\circ$ . It is also able to perform an optical zoom up to  $16\times$ . The rig is hung at the cell of a room (as shown in fig 3). On figure 4, we can see the spherical representation of the system in its initial calibrated position (that is to say  $\psi = \phi = 0$ ). On this figure we

can clearly distinguish the great epipolar circle  $C$  (in red) on the sphere  $S_p$  corresponding to the selected red point on the sphere  $S_o$ .

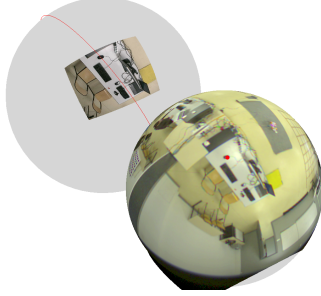


Figure 4: Spherical representation of the system after calibration

## 6.1 Scan of the epipolar line

In our simulation using PovRay the scan of the epipolar line is always very accurate since our extrinsic/intrinsic parameters are known. Figure 5 illustrates a representative experiment done with our system, it shows a sample of images acquired after selecting the region (depicted in figure 5(a)) on the fish-eye image. Here, we intentionally selected a distant and hardly distinguishable object on the omnidirectional image. This series of images is indeed the result of the computed angles steering the PTZ camera along the great epipolar circle drawn by the center of the fisheye's patch. In this sequence of images we can clearly see the epipolar line crossing the principal point (which is close to the center of the images) of every single images, furthermore this epipolar line accurately pass through the target object as expected.

## 6.2 Object detection

In our experiments we used the Harris corner detector directly on the fisheye image and on the images taken with the PTZ camera. Note that, this approach does not need any rectification of the fisheye image which is a considerable gain in computation time. Furthermore, for all object detection presented here we kept the thresholds  $\tau$  and  $\varepsilon$  fixed (0.01 and 0.1 respectively).

### 6.2.1 Experiments with synthetic images

In these synthetic experiments we point out the robustness of our algorithm in various scenarios. The metric used to quantify the quality of the object localization is the angular distance between the center of

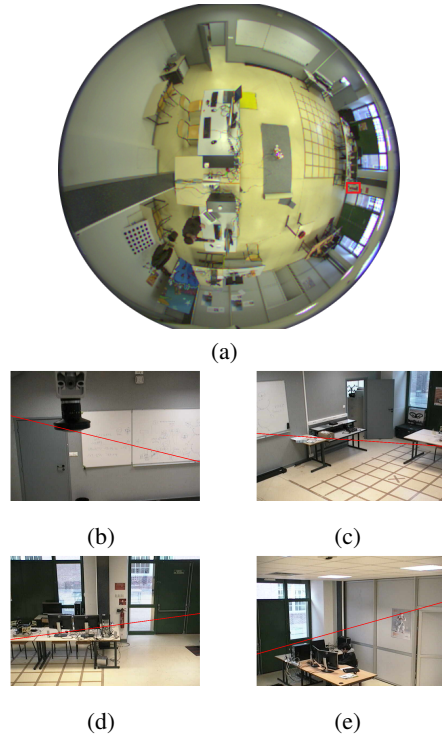


Figure 5: Scanning of an epipolar circle (a) Fisheye image with selected ROI, (b) (c) (d) (e) epipolar line through multiple images from the PTZ camera

noise variance	0	5.1	10.2	15.3	20.4	25.5	30.6	35.7
angular error (°)	6.89	7.01	3.96	5.13	5.66	5.04	4.62	6.07

Table 2: Angular error against Gaussian noise

the desired ROI with the actual position of the camera.

First of all, we tested the robustness of our algorithm against different level of pixels noise. In the experiment shown in figure 6, we have added a Gaussian noise directly on the pixel intensities -of both PTZ and fisheye images- in order to disturb the corner detection. To make the results more readable, we have displayed the resulting ROI on the non-noisy image. Table 2 depicts the angular accuracy of the method for different levels of noise. This table shows that there are no big fluctuations in the angle estimations with increase in noise up to some level. Note that even without the presence of noise we are not able to have a null error because the selected area is non-planar and cannot give a perfect matching using an homography. Nevertheless, the desired region is always in the sight of the camera whatever the level of noise. The results suggest a good robustness to this type of image noise because we do not match the corners based on their intensity.

In figure 7 we experiment our method on different surfaces, the one selected in figure 7(a) is a perfect

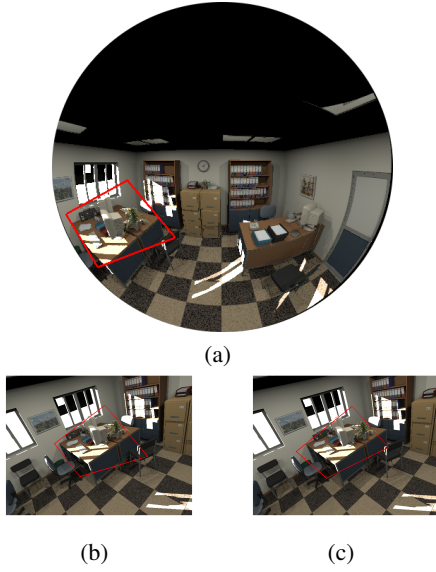


Figure 6: Results obtained against noise, (a) master image, (b), (c) results with a noise variance of 0.0 and 38.25 respectively

planar region, (b) is slightly planar while the last ROI is far from being a plane. In the first case the computed angular error is  $0.58^\circ$ ,  $0.6^\circ$  for the second and finally  $7.34^\circ$  for the last, it means that the geometry of the object has a direct incidence on the accuracy of the algorithm. However, it also shows that even in case of a totally non-planar area, our algorithm is still able to steer the camera in the right direction, keeping the desired object within the field of view of the camera. Regarding the application we are not looking for a perfect homography computation since an approximation is enough to orient the PTZ camera in the desired direction.

Figure 8 contains the results computed for the localization of the same object for different focal lengths. The angular errors obtained are in between  $2.5^\circ$  and  $4^\circ$  proving that our method performs well for various level of zoom.

### 6.2.2 Experiments with real images

The master fisheye image (figure 9(a)) used for all these experiments does not change while the images from the PTZ camera are subject to many changes in illumination or in the environment.

In the first experiment we selected a flat and well-textured surface (see the red bounding box in fig.9(a)) which is matched among images from the perspective camera. The resulting bounding box surrounding the selected area can be shown in figure 9(b), we can see that the matching is close to be perfect.

Figure 9(e) depicts a test with a planar area but with

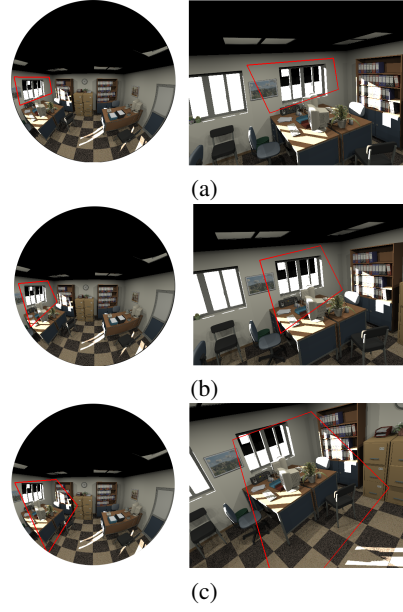


Figure 7: Test with various type of surface (a) planar (b) slightly planar (c) non-planar

a partial occlusion on the perspective image. Even in this circumstance the recognition is still very accurate. To assert our method we also applied it on non-planar regions, those experiments are depicted in figure 9(c)(d). Despite the planar patch assumption is not satisfied, the detection of the desired areas remains precise.

Figure 9(f) is the result of a detection in very challenging conditions, in fact the selected area is not flat and with occlusions. Even in these conditions our algorithm achieves very good results.

While the other images assumes a optical zoom of 1X, figure 9(g) shows a result with a zoom of 5X. The use of different zoom levels do not affect much the performances of the proposed method. Finality the last experiment (figure9(h)) proves the robustness of our method to strong illuminations changes.

## 7 Conclusion

In this paper we have presented a flexible and efficient approach to control a PTZ camera in a heterogeneous vision system. We proved that it is possible to localize a target with a PTZ camera using only information from any omnidirectional camera respecting the SVP assumption. Our method combines many advantages. Indeed it performs well even in an unknown environment by scanning an epipolar circle. Furthermore the detection of the target object is only based on geometric assumptions, while most of the



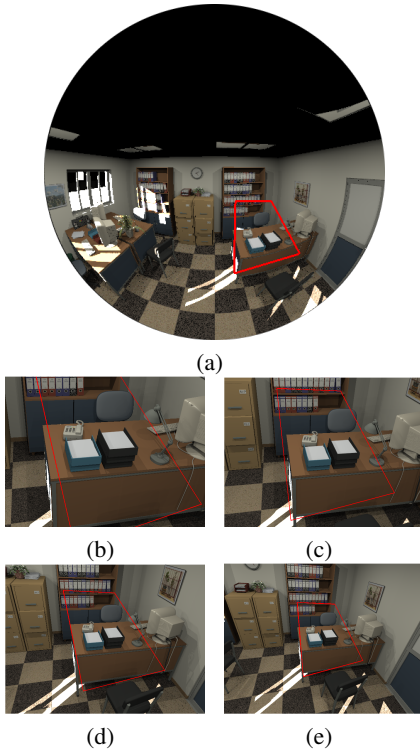


Figure 8: Detection of a ROI using different focal lengths (a) fisheye image with the selected ROI, results for an Horizontal Field Of View of (b) 40°, (c) 50°, (d) 65°, (e) 80°

hybrid matching methods in the literature use photometric descriptors. Consequently, our approach can deal with strong distortions, illumination changes and big scale difference without any rectification of the omnidirectional image. It is also important to note that this template matching can be used for any calibrated hybrid vision system.

The provided set of qualitative and quantitative results show the great accuracy and flexibility offers by our method which is capable to detect any kind of ROI. This work can be useful in many robotics or video surveillance applications, for instance as an initialization for collaborative object tracking.

## ACKNOWLEDGEMENTS

This work was supported by DGA (Direction Generale de l'Armement) and the regional council of Burgundy.

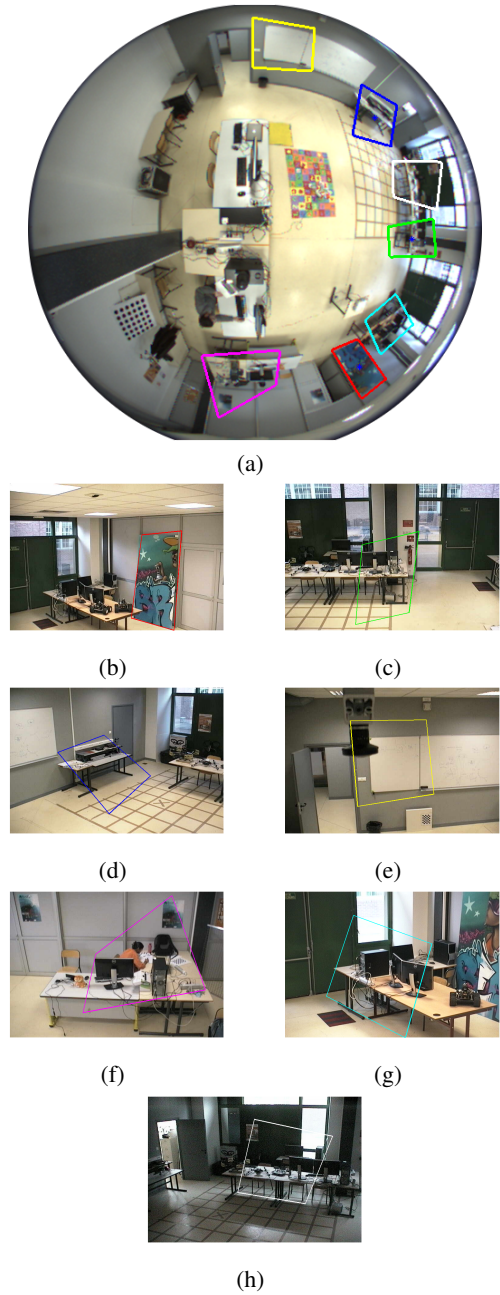


Figure 9: Detection of various objects (a) Fisheye image with selected ROIs, (b) (c) (d) (e) (f) (g) and (h) Detected targets through PTZ images

## REFERENCES

- Adorni, G., Bolognini, L., Cagnoni, S., and Mordonini, M. (2002). Stereo obstacle detection method for a hybrid omni-directional/pin-hole vision system. In *RoboCup 2001: Robot Soccer World Cup V*, pages 244–250, London, UK. Springer-Verlag.
- Agapito, L., Hayman, E., and Reid, I. (2001). Self-

- calibration of rotating and zooming cameras. *International Journal of Computer Vision*, 45(2):107–127.
- Amine Iraqui, H., Dupuis, Y., Bouteau, R., Ertaud, J.-Y., and Savatier, X. (2010). Fusion of omnidirectional and ptz cameras for face detection and tracking. In *Emerging Security Technologies (EST), 2010 International Conference on*, pages 18–23. IEEE.
- Badri, J., Tilmant, C., Lavest, J., Pham, Q., and Sayd, P. (2007). Camera-to-camera mapping for hybrid pan-tilt-zoom sensors calibration. In *SCIA*, pages 132–141.
- Barreto, J. P. and Araujo, H. (2001). Issues on the geometry of central catadioptric image formation. In *CVPR*, pages 422–427.
- Bazin, J., Kim, S., Ghoi, D., J.Y.Lee, and Kweon, I. (2011). Mixing collaborative and hybrid vision devices for robotics applications. *Journal of Korea Robotics Society*.
- Bouguet, J. Y. (2008). Camera calibration toolbox for Matlab.
- Chen, C.-H., Yao, Y., Page, D. L., Abidi, B. R., Koschan, A., and Abidi, M. A. (2008). Heterogeneous fusion of omnidirectional and ptz cameras for multiple object tracking. *IEEE Trans. Circuits Syst. Video Techn.*, 18(8):1052–1063.
- Courbon, J., Y.Mezouar, and Martinet, P. (2012). Evaluation of the unified model of the sphere for fisheye cameras in robotic applications. *Advanced Robotics*, 26(8-9):947–967.
- Cui, Y., Samarasekera, S., Huang, Q., Greinhagen, M., and Enhagen, M. G. (1998). Indoor monitoring via the collaboration between a peripheral sensor and a foveal sensor. In *In Proc. of the IEEE Workshop on Visual Surveillance*, pages 2–9. IEEE Computer Society.
- Cyganek, B. and Gruszczyński, S. (2013). Hybrid computer vision system for drivers’ eye recognition and fatigue monitoring. *Neurocomputing*.
- Ding, C., Song, B., Morye, A., Farrell, J. A., and Roy-Chowdhury, A. K. (2012). Collaborative sensing in a distributed ptz camera network. *Image Processing, IEEE Transactions on*, 21(7):3282–3295.
- Eynard, D., Vasseur, P., Demonceaux, C., and Frémont, V. (2012). Real time uav altitude, attitude and motion estimation from hybrid stereovision. *Autonomous Robots*, 33(1-2):157–172.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395.
- Fujiki, J., Torii, A., and Akaho, S. (2007). Epipolar geometry via rectification of spherical images. In *Proceedings of the 3rd international conference on Computer vision/computer graphics collaboration techniques, MIRAGE’07*, pages 461–471, Berlin, Heidelberg. Springer-Verlag.
- Gould, S., Arfvidsson, J., Kaehler, A., Messner, M., Bradski, G., Baumstarck, P., Chung, S., and Ng, A. Y. (2007). Peripheral-foveal vision for real-time object recognition and tracking in video. In *In International Joint Conference on Artificial Intelligence (IJCAI)*.
- Liao, H. C. and Cho, Y. C. (2008). A new calibration method and its application for the cooperation of wide-angle and pan-tilt-zoom cameras. *Information Technology Journal*, 7(8):1096–1105.
- Marr, D. and Poggio, T. (1977). A theory of human stereo vision. Technical report, Cambridge, MA, USA.
- Mei, C., Benhimane, S., Malis, E., and Rives, P. (2008). Efficient homography-based tracking and 3-d reconstruction for single-viewpoint sensors. *IEEE Transactions on Robotics*, 24(6):1352–1364.
- Mei, C. and Rives, P. (2007). Single view point omnidirectional camera calibration from planar grids. In *IEEE International Conference on Robotics and Automation*.
- Micheloni, C., Rinner, B., and Foresti, G. L. (2010). Video analysis in pan-tilt-zoom camera networks. *Signal Processing Magazine, IEEE*, 27(5):78–90.
- Neves, A. J., Martins, D. A., and Pinho, A. J. (2008). A hybrid vision system for soccer robots using radial search lines. In *Proc. of the 8th Conference on Autonomous Robot Systems and Competitions, Portuguese Robotics Open-ROBOTICA*, pages 51–55.
- Puwein, J., Ziegler, R., Ballan, L., and Pollefeys, M. (2012). Ptz camera network calibration from moving people in sports broadcasts. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pages 25–32. IEEE.
- Raj, A., Khemmar, R., Eratud, J. Y., and Savatier, X. (2013). Face detection and recognition under heterogeneous database based on fusion of catadioptric and ptz vision sensors. In *Proceedings of the 8th International Conference on Computer Recognition Systems CORES 2013*, pages 171–185. Springer.
- Rameau, F., Habed, A., Demonceaux, C., Sidibé, D., and Fofi, D. (2012). Self-calibration of a ptz camera using new lmi constraints. In *ACCV*.
- Scotti, G., Marcenaro, L., Coelho, C., Selvaggi, F., and Regazzoni, C. (2005). Dual camera intelligent sensor for high definition 360 degrees surveillance. *IEEE Proceedings-Vision, Image and Signal Processing*, 152(2):250–257.
- Ying, X. and Hu, Z. (2004). Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. In *ECCV*, pages 442–455.